

Linear quantile mixed models

Marco Geraci · Matteo Bottai

Received: 7 September 2012 / Accepted: 24 January 2013 / Published online: 20 February 2013
© Springer Science+Business Media New York 2013

Abstract Dependent data arise in many studies. Frequently adopted sampling designs, such as cluster, multilevel, spatial, and repeated measures, may induce this dependence, which the analysis of the data needs to take into due account. In a previous publication (Geraci and Bottai in *Biostatistics* 8:140–154, 2007), we proposed a conditional quantile regression model for continuous responses where subject-specific random intercepts were included to account for within-subject dependence in the context of longitudinal data analysis. The approach hinged upon the link existing between the minimization of weighted absolute deviations, typically used in quantile regression, and the maximization of a Laplace likelihood. Here, we consider an extension of those models to more complex dependence structures in the data, which are modeled by including multiple random effects in the linear conditional quantile functions. We also discuss estimation strategies to reduce the computational burden and inefficiency associated with the Monte Carlo EM algorithm we have proposed previously. In particular, the estimation of the fixed regression coefficients and of the random effects' covariance matrix is based on a combination of Gaussian quadrature approximations and non-smooth optimization algorithms. Finally, a simulation study and a number of applications of our models are presented.

Keywords Best linear predictor · Clarke's derivative · Hierarchical models · Gaussian quadrature

1 Introduction

Conditional quantile regression (QR) pertains to the estimation of unknown quantiles of an outcome as a function of a set of covariates and a vector of fixed regression coefficients. Generally, QR estimation makes no assumption on the shape of the distribution of the outcome (Bosovich 1757; Wagner 1959; Barrodale and Roberts 1978; Bassett and Koenker 1978; Koenker and Bassett 1978). Its ability to provide a thorough description of the distributional effects has contributed to make QR attractive in several fields. See for example Yu et al. (2003) and Koenker (2005) for an overview of recent applications.

A number of sampling designs such as multilevel, longitudinal and cluster sampling typically require the application of statistical methods that allow for the correlation between observations that belong to the same unit or cluster. Mixed effects models, also known as multilevel or hierarchical or random-effects models, represent highly popular and flexible models to analyze complex data. They model and estimate between-cluster variability by means of cluster-specific random effects. These, in turn, provide a modeling structure for estimating the intraclass correlation coefficient (ICC). For example, let Z be a block-diagonal matrix with diagonal blocks given by vectors of ones and I be an identity matrix of appropriate dimensions. In the two-level random intercept model

$$y|u = \mu + Zu + \varepsilon, \\ u \sim N(0, \psi_u^2 I), \varepsilon \sim N(0, \psi^2 I), u \perp \varepsilon, \quad (1)$$

M. Geraci (✉)
Centre for Paediatric Epidemiology and Biostatistics,
MRC Centre of Epidemiology for Child Health,
Institute of Child Health, University College London,
30 Guilford Street, London WC1N 1EH, UK
e-mail: m.geraci@ucl.ac.uk

M. Bottai
Unit of Biostatistics, Institute of Environmental Medicine,
Karolinska Institutet, Nobels väg 13, 17177 Stockholm, Sweden

the ratio of the variance of the random intercepts to the total variance defines the proportion of variance explained by clustering, namely $ICC = \psi_u^2 / (\psi_u^2 + \psi^2)$. Inference can be then conducted at the population (marginal) or cluster (conditional) level. Indeed, this represents an advantage of random-effects models over strictly marginal models for the latter cannot provide conditional inferences (Lee and Nelder 2004). In these models, both fixed and random effects are assumed to be purely location-shift effects.

In the last few years, the need for extending QR for independent data to clustered data has led to several and quite distinct approaches. These can be roughly classified into two groups: distribution-free and likelihood-based. The former includes fixed effects (Koenker 2004; Lamarche 2010; Galvao and Montes-Rojas 2010; Galvao 2011) and weighted (Lipsitz et al. 1997; Karlsson 2008; Fu and Wang 2012) approaches. The latter are mainly based on the asymmetric Laplace (AL) density (Geraci and Bottai 2007; Liu and Bottai 2009; Yuan and Yin 2010; Lee and Neocleous 2010; Farcomeni 2012) or other parametric distributions (Reich et al. 2010a).

These categories are by no means mutually exclusive. For example, penalty methods as those proposed by Koenker (2004) might have a strict relationship with the asymmetric Laplace regression with double-exponential random effects as suggested by Geraci and Bottai (2007). Yet, this has not been fully explored—see Lamarche (2010) for a discussion on the asymptotics of L_1 penalized fixed-intercept models. Nor the classification is exhaustive. In the approach by Canay (2011) location-shift fixed effects are eliminated by a transformation. Also, other approaches might involve modeling the moments of a distribution function using a parametric family (e.g., Rigby and Stasinopoulos 2005) and deriving the quantiles of the response by inversion. Yet, the spirit of a likelihood-based approach to QR is different from a probability model fitting exercise, where a given distribution is assumed to be the ‘true’ distribution.

In Sect. 2, we briefly review Geraci and Bottai’s AL-based approach (Geraci and Bottai 2007, hereafter GB) and we introduce a generalization of the model proposed therein. In Sect. 3, we describe an estimation process based on numerical integration and nonsmooth optimization. Inferential issues are also discussed. A simulation study is presented in Sect. 4. We then conclude with some applications (Sect. 5) and final remarks (Sect. 6). All computations were performed using the package `lqmm` (Geraci 2012) for the statistical programming environment R (R Development Core Team 2012).

2 Linear quantile mixed models

2.1 Random-effects models with asymmetric Laplace error

A continuous random variable $w \in \mathbb{R}$ is said to follow an asymmetric Laplace density with parameters (μ, σ, τ) , $w \sim \text{AL}(\mu, \sigma, \tau)$, if its density can be expressed as

$$p(w|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\frac{1}{\sigma} \rho_\tau(w - \mu)\right\},$$

where $-\infty < \mu < +\infty$ is the location parameter, $\sigma > 0$ is the scale parameter, $0 < \tau < 1$ is the skew parameter, and $\rho_\tau(v) = v\{\tau - I(v < 0)\}$ is the loss function, with $I(\cdot)$ denoting the indicator function. Note that ρ_τ is the asymmetrically weighted L_1 function used in quantile regression problems (Koenker and Bassett 1978). The AL distribution first appeared in Hinkley and Revankar (1977) under a different parameterization. See Yu and Zhang (2005) for more details on this distribution. In our case, the location parameter μ is of great interest as this is the τ th quantile of w , i.e. $\Pr(w \leq \mu) = \tau$.

Let $y \in \mathbb{R}$ be a response variable with unknown, continuous distribution function F_y . For a fixed skewness parameter τ , it is convenient to estimate the τ th quantile of y , $\mu^{(\tau)}$, from the model

$$y = \mu^{(\tau)} + \varepsilon^{(\tau)},$$

where $\varepsilon^{(\tau)} \sim \text{AL}(0, \sigma, \tau)$. Note that this assumption is ancillary as we do not assume that F_y is AL.

For a random variable w composed of n independent variates w_i with common skew and scale parameters, but generic location $\mu = (\mu_1, \dots, \mu_n)'$, $w_i \sim \text{AL}(\mu_i, \sigma, \tau)$, $i = 1, \dots, n$, we use the simplified notation

$$p(w|\mu, \sigma, \tau) = \sigma_n(\tau) \exp\left\{-\frac{1}{\sigma} \rho_\tau(w - \mu)\right\},$$

where $\sigma_n(\tau) = \tau^n(1 - \tau)^n/\sigma^n$ and $\rho_\tau(y - \mu) = \sum_{i=1}^n \rho_\tau(w_i - \mu_i)$.

In GB, we proposed a random-intercepts QR model for longitudinal data using the AL to model the τ th conditional quantile of a continuous response variable. In particular, we assumed the following regression function:

$$Q_{y|u}(\tau|x, u) = X\beta^{(\tau)} + u,$$

where (y, X) represents longitudinal data, u a vector of subject-specific random effects and $Q_{y|u}$ denotes the inverse of the (unknown) distribution function $F_{y|u}$. The τ th regression quantile of $y|u$ was then estimated under the convenient assumption $y|u \sim \text{AL}(X\beta^{(\tau)} + u, \sigma^{(\tau)}, \tau)$, where the τ -dependent parameters $\beta^{(\tau)}$ and $\sigma^{(\tau)}$ have a frequentist interpretation. A Bayesian character to this model, however, has been attributed by others (Reich et al. 2010b;

Li et al. 2010). The link between the L_1 norm regression problem and the AL-based estimation of the coefficients $\beta^{(\tau)}$ has been discussed by Koenker and Machado (1999) and Yu and Moyeed (2001). Here, we emphasize that the mode of $y|u$ is $X\beta^{(\tau)} + u$, for any $\tau \in (0, 1)$.

A model with random intercepts only cannot obviously account for between-clusters heterogeneity associated with given explanatory variables. This is the case, for example, of growth curves in which both intercepts and slopes of temporal trajectories differ by subject due to genetic and environmental effects. In the next section, we extend our model to include multiple random effects.

2.2 Model generalization

Consider clustered data in the form $(x'_{ij}, z'_{ij}, y_{ij})$, for $j = 1, \dots, n_i$ and $i = 1, \dots, M$, $N = \sum_i n_i$, where x'_{ij} is the j th row of a known $n_i \times p$ matrix X_i , z'_{ij} is the j th row of a known $n_i \times q$ matrix Z_i and y_{ij} is the j th observation of the response variable in the i th cluster. Mixed effects models represent a common and well-known class of regression models used to analyze data coming from similar designs. A typical formulation of a linear mixed model (LMM) for clustered data is

$$y_{ij} = x'_{ij}\beta + z'_{ij}u_i + \varepsilon_{ij}, \quad j = 1, \dots, n_i, \quad i = 1, \dots, M,$$

where β and u_i , $i = 1, \dots, M$, are, respectively, fixed and random effects associated with p and q model covariates and y is assumed to follow a multivariate normal distribution characterized by some parameter θ . Within this framework, the target of the analysis is the mean of the response, whether conditional, i.e. $E(y_i|u)$, or marginal, i.e. $E(y_i)$, will depend on the purpose of the analysis. The likelihood of the LMM specified above is, from a modeling standpoint, conditional. The term marginal refers to the likelihood integrated with respect to u . Throughout the paper, we will use such distinction.

There are several reasons why an analyst might consider a random-effects approach for their data while wanting to go beyond a location-shift model. A quantile regression approach allows for departures from location-shift assumptions, typical of LMMs, in which covariates do not affect the shape of the distribution. Here, we follow the same approach of GB but we provide a framework within which modeling of the regression quantiles of clustered continuous outcomes is more general and estimation is computationally more efficient.

We introduce the convenient assumption that the $y_i = (y_{i1}, \dots, y_{in_i})'$, $i = 1, \dots, M$, conditionally on a $q \times 1$ vector of random effects u_i , are independently distributed according to a joint AL with location and scale parameters given by $\mu_i^{(\tau)} = X_i\theta_x^{(\tau)} + Z_iu_i$ and $\sigma^{(\tau)}$, where $\theta_x^{(\tau)} \in \mathbb{R}^p$

is a vector of unknown fixed effects. The skew parameter τ is set *a priori* and defines the quantile level to be estimated. Also, we assume that $u_i = (u_{i1}, \dots, u_{iq})'$, for $i = 1, \dots, M$, is a random vector independent from the model's error term and distributed according to $p(u_i|\Psi^{(\tau)})$, where $\Psi^{(\tau)}$ is a $q \times q$ covariance matrix. Note that all the parameters are τ -dependent. The random effects vector u depends on τ through $\Psi^{(\tau)}$. We will omit the superscript τ when this is not source of confusion. Moreover, we assume that the random effects are zero-median vectors. Other scenarios may include the case in which u are not zero-centered (see for example Koenker 2005, p. 281) and/or are not symmetric. A discussion on proposed distributions for the random effects is deferred to the next section.

If we let $u = (u'_1, \dots, u'_M)'$, $y = (y'_1, \dots, y'_M)'$, $X = [X'_1 | \dots | X'_M]'$ and $Z = \bigoplus_{i=1}^M Z_i$, $\mu^{(\tau)} = X\theta_x^{(\tau)} + Zu$, the joint density of (y, u) based on M clusters for the τ th linear quantile mixed model (LQMM) is given by

$$\begin{aligned} p(y, u|\theta_x^{(\tau)}, \sigma^{(\tau)}, \Psi^{(\tau)}) \\ &= p(y|\theta_x^{(\tau)}, \sigma^{(\tau)}, u)p(u|\Psi^{(\tau)}) \\ &= \prod_{i=1}^M p(y_i|\theta_x^{(\tau)}, \sigma^{(\tau)}, u_i)p(u_i|\Psi^{(\tau)}). \end{aligned} \quad (2)$$

Throughout the paper we will assume that $\Psi^{(\tau)} \in S_{++}^q$, where S_{++}^q is the set of real symmetric positive-definite $q \times q$ matrices. The likelihood in (2) can be equivalently seen as arising from the model

$$y = \mu^{(\tau)} + \varepsilon^{(\tau)},$$

with independent and identically distributed errors $\varepsilon_{ij}^{(\tau)} \sim \text{AL}(0, \sigma, \tau)$.

The formulation of the model in (2) is general. For $Z_i = (1, \dots, 1)'$, $i = 1, \dots, M$, the LQMM corresponds to the random-intercepts model developed by GB. We obtain the i th contribution to the marginal likelihood by integrating out the random effects, which leads to

$$L_i(\theta_x, \sigma, \Psi|y_i) = \int_{R^q} p(y_i, u_i|\theta_x, \sigma, \Psi) du_i, \quad (3)$$

where R^q denotes the q -dimensional Euclidean space. We denote the marginal log-likelihood by $\ell_i(\theta_x, \sigma, \Psi|y) = \log L_i(\theta_x, \sigma, \Psi|y)$, $i = 1, \dots, M$.

Clearly, since inference in model (2) is done separately for each τ , quantile crossing phenomena may arise within the convex hull of X , perhaps as a result of model misspecification (Koenker 2005, p. 55). As explained by Lum and Gelfand (2012), the stronger constraint of a joint quantile specification is replaced by the weaker stochastic ordering of the asymmetric Laplace. *Ad-hoc* solutions to avoid quantile-crossing have been proposed (He 1997; Zhao 2000).

3 Estimation

GB proposed to fit the random-intercepts QR model by using a Monte Carlo EM algorithm. This approach on the one hand avoids evaluating a multidimensional integral, but on the other hand, it can be computationally demanding. In a simulation study, Alhamzawi et al. (2011) compared Gibbs sampling with GB's EM algorithm and with Yuan and Yin's (2010) MCMC algorithm. In terms of relative bias, the three approaches were found to be analogous. However, the EM algorithm showed poorer computational efficiency. Modeling complex structures of the random effects increases the computational requirements. In this section, we will explore alternative strategies based on different optimization techniques.

Our goal is to evaluate the integral

$$p(y_i|\theta_x, \sigma, \Psi) = \sigma_{n_i}(\tau) \int_{\mathbb{R}^q} \exp\left\{-\frac{1}{\sigma} \rho_\tau(y_i - \mu_i)\right\} p(u_i|\Psi) du_i. \quad (4)$$

The marginal density in (4) can be characterized by using Theorem 6 by Prékopa (1973).

Theorem 1 (Prékopa 1973) *Let $f(x, y)$ be a function of $p + q$ variables where x is a p -component and y is a q -component vector. Suppose that f is logarithmic concave in \mathbb{R}^{p+q} and let A be a convex subset of \mathbb{R}^q . Then the function of the variable x*

$$g(x) = \int_A f(x, y) dy$$

is logarithmic concave in the entire space \mathbb{R}^p .

If $p(u_i|\Psi)$ is log-concave in u , the integrand in (4) will be log-concave in y and Prékopa's theorem will apply to $p(y_i|\theta_x, \sigma, \Psi)$. As we shall see, the distributions of the random effects considered in this study are log-concave. The joint log-likelihood, however, is not a concave function of the scale and variance parameters for which a reparameterization is required in order to apply the above results.

3.1 Numerical integration

The integral in (4) has the form $\int_{\mathbb{R}^q} f(u)w(u) \prod_q du_q$. By choosing a suitable weighting function or kernel $w(u)$ and upon a change of the integration variable where necessary, the q -dimensional Gaussian quadrature formula, based on q successive applications of simple one-dimensional rules, provides the approximation

$$\int_{\mathbb{R}^q} f(u)w(u) du \approx \sum_{k_1=1}^K \cdots \sum_{k_q=1}^K f(v_{k_1, \dots, k_q}) \prod_{l=1}^q w_{k_l},$$

where K is a given integer. The abscissas $v_{k_1, \dots, k_q} = (v_{k_1}, \dots, v_{k_q})'$ and the weights w_{k_l} , $k_l = 1, \dots, K$, $l = 1, \dots, q$, are chosen so that the approximation is exact if $f(u)$ is a polynomial of a given total order. More precisely, the product rule defined above would be exact for a tensor product of univariate polynomials.

For this reason, the product rule entails a 'curse of dimensionality', an exponential increase of the number of evaluations of the integrand function. For example, we would need 3,200,000 function evaluations for a 5-dimensional quadrature rule based on 20 nodes and 64,000,000 for adding only one random effect, let alone the total number of evaluations necessary to convergence if the parameter's estimation algorithm is iterative (as it will be seen to be the case). Fortunately, a satisfactory accuracy is typically achieved with less than 20 nodes.

The choice of an appropriate distribution for the random effects is not straightforward. As recognized by GB, robustness issues apply to the distribution of $y|u$ as well as of u . As a robust alternative to the Gaussian choice, we suggested the use of the symmetric Laplace. This choice led to a regression model which, after a rescaling (Geraci and Bottai 2007, p. 146), was similar to the penalized model proposed by Koenker (2004). As noted by Koenker (2004), using L_1 penalties in fixed-effect models is convenient as this produces an elegant form of the penalized objective function apt to be solved via linear programming algorithms. See also Koenker and Mizera (2004) for quantile bivariate smoothing using L_1 penalties.

In the following, we will focus explicitly on two types of distributions, namely Gaussian and Laplacian. It is immediate to verify that these assumptions correspond to applying, respectively, a Gauss–Hermite and a Gauss–Laguerre quadrature to the integral in (4). More general considerations can be made with regard to the use of symmetric as well as asymmetric kernels belonging to the exponential family.

3.1.1 Normal random effects

Under the assumption of normal random effects, the approximation of the integral in (4) by Gauss–Hermite quadrature is, bar a proportionality constant,

$$\begin{aligned} & \int_{\mathbb{R}^q} \sigma^{-n_i} |\Psi|^{-1/2} \exp\left\{-\frac{1}{\sigma} \rho_\tau(y_i - X_i \theta_x - Z_i u_i)\right\} \\ & \quad \times \exp(-u_i' \Psi^{-1} u_i / 2) du_i \\ & = \int_{\mathbb{R}^q} \exp\left\{-\frac{1}{\sigma} \rho_\tau[y_i - X_i \theta_x - Z_i (\Psi')^{1/2} v] - n_i \log \sigma\right\} \\ & \quad \times \exp(-\|v\|^2 / 2) dv \end{aligned}$$

$$\simeq \sum_{k_1=1}^K \cdots \sum_{k_q=1}^K \exp \left\{ -\frac{1}{\sigma} \rho_{\tau} [y_i - X_i \theta_x - Z_i (\Psi')^{1/2} v_{k_1, \dots, k_q}] - n_i \log \sigma \right\} \times \prod_{l=1}^q w_{k_l},$$

with nodes $v_{k_1, \dots, k_q} = (v_{k_1}, \dots, v_{k_q})'$ and weights w_{k_l} , $l = 1, \dots, q$.

The (marginal) log-likelihood for all clusters is approximated by

$$\begin{aligned} \ell_{\text{app}}(\theta_x, \sigma, \Psi | y) &= \sum_i^M \log \left\{ \sum_{k_1=1}^K \cdots \sum_{k_q=1}^K p(y_i | \theta_x, \sigma, (\Psi')^{1/2} v_{k_1, \dots, k_q}) \right. \\ &\quad \left. \times \prod_{l=1}^q w_{k_l} \right\}. \end{aligned} \quad (5)$$

The integral above can be recognized as a normal-Laplace convolution (Reed 2006). This type of distribution is known in a special form in meta-analysis (Demidenko 2004).

3.1.2 Robust random effects

We consider independent random effects under the assumption of a symmetric Laplace distribution, therefore $\Psi = \bigoplus_{l=1}^q \psi_l$. Since each one-dimensional integral in (4) can be split around zero, the Gauss–Laguerre quadrature can be applied to the interval $[0, \infty)$ and, by symmetry, to the interval $(-\infty, 0]$. This results in the following approximation:

$$\begin{aligned} &\int_{R^q} \sigma^{-n_i} \left(\prod_{l=1}^q \psi_l^{-1} \right) \exp \left\{ -\frac{1}{\sigma} \rho_{\tau} (y_i - X_i \theta_x - Z_i u_i) \right\} \\ &\quad \times \exp \left(-\sum_{l=1}^q |u_{il}|/2 \psi_l \right) du_i \\ &= \int_{R^q} \exp \left\{ -\frac{1}{\sigma} \rho_{\tau} (y_i - X_i \theta_x - Z_i \Psi' v) - n_i \log \sigma \right\} \\ &\quad \times \exp \left(-\sum_{l=1}^q |v_l|/2 \right) dv \\ &\simeq \sum_{k_1=1}^K \cdots \sum_{k_q=1}^K \exp \left\{ -\frac{1}{\sigma} \rho_{\tau} (y_i - X_i \theta_x - Z_i \Psi' v_{k_1, \dots, k_q}) \right. \\ &\quad \left. - n_i \log \sigma \right\} \prod_{l=1}^q w_{k_l}, \end{aligned}$$

with nodes $v_{k_1, \dots, k_q} = (v_{k_1}, \dots, v_{k_q})'$ and weights w_{k_l} , $l = 1, \dots, q$, opportunely chosen.

The log-likelihood for all clusters is approximated by

$$\begin{aligned} \ell_{\text{app}}(\theta_x, \sigma, \Psi | y) &= \sum_i^M \log \left\{ \sum_{k_1=1}^K \cdots \sum_{k_q=1}^K p(y_i | \theta_x, \sigma, \Psi' v_{k_1, \dots, k_q}) \prod_{l=1}^q w_{k_l} \right\}. \end{aligned} \quad (6)$$

There are several papers on variations of the univariate Laplace distribution (Kozubowski and Nadarajah 2010). A valid multivariate extension of the asymmetric Laplace distribution that imposes a general covariance structure is not obvious. For example, Kotz et al. (2000) proposed a multivariate asymmetric Laplace distribution whose density in the n -variate symmetric case is given by

$$\begin{aligned} p(y | \Sigma) &= 2(2\pi)^{-n/2} |\Sigma|^{-1/2} (y' \Sigma^{-1} y / 2)^{\lambda/2} \\ &\quad \times K_{\lambda}(\sqrt{2y' \Sigma^{-1} y}), \end{aligned} \quad (7)$$

with $\text{Cov}(y) = \Sigma$, $\lambda = 1 - n/2$ and $K_{\lambda}(t)$ is the modified Bessel function of the third kind. For $n = 1$ and $\Sigma = \sigma^2$, the density in (7) reduces to

$$p(y | \sigma) = \frac{\sqrt{2}}{2\sigma} \exp \left(-\frac{\sqrt{2}}{\sigma} |y| \right),$$

that is $y \sim \text{AL}(0, \frac{\sigma}{2\sqrt{2}}, \frac{1}{2})$. It can be shown that for a $n \times n$ real matrix G , $\text{Cov}(Gy) = G \Sigma G'$ (Kotz et al. 2000).

The use of (7) for correlated random effects is uncertain. Even though it is easy to re-scale Ψ to a diagonal matrix, the joint density does not factorize into q AL variates (Eltoft et al. 2006) and, therefore, the q -dimensional quadrature can not be based on q successive applications of one-dimensional rules. Therefore, at the moment, we do not consider any of the available proposals for a generalized Laplace distribution as suitable for our purposes. See for example Liu and Bottai (2009) for some results on the use of Kotz et al.'s (2000) multivariate Laplace distribution.

3.2 Nonsmooth optimization

The nondifferentiability of the loss function ρ_{τ} at points where $y_{ij} - x'_{ij} \theta_x - z'_{ij} u = 0$ interferes with the standard theory of smooth optimization. Subgradient optimization and derivative-free optimization techniques (e.g., coordinate and pattern-search methods, modified Nelder–Mead methods, implicit filtering) have been developed to tackle non-standard optimization problems. Here, we briefly study the Clarke's derivative of the loss function as starting point for a nonsmooth optimization approach. In his book, first published in 1983, Clarke (1990) developed a general theory of nonsmooth analysis that leads to a powerful and elegant approach to mathematical programming. Here, we focus our attention on theorems for Lipschitz functions which play an important role in Clarke's treatise.

We begin by characterizing the (approximated) marginal likelihood as a Lipschitz function. The covariance matrix of the random effects is reparameterized in terms of an m -dimensional vector of non-redundant parameters θ_z , i.e. $\Psi(\theta_z)$. Let us define $\theta = (\theta'_x, \theta'_z)'$ and consider the i th contribution to the likelihood in (5) or (6), rewritten compactly as

$$\ell_{\text{app},i}(\theta, \sigma)$$

$$= \log \sigma_{n_i}(\tau) + \log \sum_{\mathbf{k}} \exp \left\{ -\frac{1}{\sigma} \rho_{\tau}(y_i - \tilde{X}_{i,\mathbf{k}}\theta) \right\} \prod_{l=1}^q w_{k_l}, \quad (8)$$

where $\tilde{X}_{i,\mathbf{k}} = [X_i | (v'_k \otimes Z_i) T_q]$ has row vectors $\tilde{x}'_{ij,\mathbf{k}}$, $v_{\mathbf{k}}$ is a $q \times 1$ vector of nodes and $\mathbf{k} = (k_1, \dots, k_q)'$. T_q is a matrix of order $q^2 \times m$ so that $\text{vec}(\Psi^{1/2}) = T_q \theta_z$ (Gauss–Hermite) or $\text{vec}(\Psi) = T_q \theta_z$ (Gauss–Laguerre). Since the likelihood function is strictly differentiable with respect to σ , we focus on θ alone for brevity.

It can be noted that $\ell_{\text{app},i}$ is a real-valued function of θ given by the composition $g \circ h$, where $h: \mathbb{R}^{p+m} \rightarrow \mathbb{R}^{n_i K^q}$ and $g: \mathbb{R}^{n_i K^q} \rightarrow \mathbb{R}$. Each component function $h_{ij,\mathbf{k}}$ of h ,

$$h_{ij,\mathbf{k}}(\theta) = \rho_{\tau}(y_{ij} - \tilde{x}'_{ij,\mathbf{k}}\theta)/\sigma,$$

is Lipschitz near θ , so is

$$g(h(\theta)) = \log \sigma_{n_i}(\tau) + \log \sum_{\mathbf{k}} \exp \left\{ -\sum_j h_{ij,\mathbf{k}}(\theta) \right\} \prod_{l=1}^q w_{k_l}$$

near $h(\theta)$. This implies that $g \circ h$ is Lipschitz near θ .

Then, we calculate the generalized gradient for Lipschitz functions (Clarke 1990). We have that (i) $-g$ is regular at $h(\theta)$, (ii) each $h_{ij,\mathbf{k}}$ is regular at θ , and (iii) every $n_i K^q$ -dimensional element λ of $\partial(-g)(h(\theta))$ has nonnegative components

$$\begin{aligned} \lambda_{ir,\mathbf{h}} &\equiv \partial(-g)(h(\theta))_{ir,\mathbf{h}} \\ &= \frac{\exp\{-h_{ir,\mathbf{h}}(\theta) - \sum_{j \neq r} h_{ij,\mathbf{h}}(\theta)\} \prod_{l=1}^q w_{h_l}}{\sum_{\mathbf{k}} \exp\{-\sum_j h_{ij,\mathbf{k}}(\theta)\} \prod_{l=1}^q w_{k_l}}, \end{aligned}$$

for $r = 1, \dots, n_i$ and $\mathbf{h} = (h_1, \dots, h_q)'$, $h_l = 1, \dots, K$, $l = 1, \dots, q$.

For the chain rule I (Clarke 1990, Theorem 2.3.9, p. 42), it follows that

$$\begin{aligned} \partial(-g \circ h)(\theta) &= \mathcal{C} \left\{ \sum \lambda_{ij,\mathbf{k}} \xi_{ij,\mathbf{k}} : \xi_{ij,\mathbf{k}} \in \partial h_{ij,\mathbf{k}}(\theta), \right. \\ &\quad \left. \lambda \in \partial(-g)(h(\theta)) \right\}, \quad (9) \end{aligned}$$

where the summation is extended to all j 's and \mathbf{k} 's and \mathcal{C} denotes the weak*-closed convex hull. For the chain rule II

(Clarke 1990, Theorem 2.3.10, p. 45)

$$\xi_{ij,\mathbf{h}} \equiv \partial h_{ij,\mathbf{h}}(\theta) = \begin{cases} -\frac{\tau}{\sigma} \tilde{x}_{ij,\mathbf{h}}, & \text{if } y_{ij} - \tilde{x}'_{ij,\mathbf{h}}\theta > 0, \\ -\frac{(\tau-1)}{\sigma} \tilde{x}_{ij,\mathbf{h}}, & \text{if } y_{ij} - \tilde{x}'_{ij,\mathbf{h}}\theta < 0, \\ -\frac{1}{\sigma}(\omega + \tau - \frac{1}{2}) \tilde{x}_{ij,\mathbf{h}} : |\omega| \leq \frac{1}{2}, & \\ & \text{if } y_{ij} - \tilde{x}'_{ij,\mathbf{h}}\theta = 0. \end{cases}$$

For estimation purposes, the contribution to the gradient of the zero-probability observation $y_{ij} = \tilde{x}'_{ij,\mathbf{h}}\theta$ is set at $-\frac{\tau}{\sigma} \tilde{x}_{ij,\mathbf{h}}$ for all i, j and \mathbf{h} , that is $\omega = \frac{1}{2}$.

If a local optimum is attained at $\hat{\theta}$, solution to the minimization problem

$$\min_{\theta} \left\{ -\sum_{i=1}^M \ell_{\text{app},i}(\theta, \sigma) \mid \theta_x \in \mathbb{R}^p, \Psi(\theta_z) \in S_{++}^q \right\},$$

then $0 \in \partial -\sum_i \ell_{\text{app},i}$ (Clarke 1990). In practice, the constraint $\Psi(\theta_z) \in S_{++}^q$ can be imposed *a posteriori* by calculating the nearest symmetric positive definite matrix (Higham 2002) after θ_z is estimated or, *a priori*, by a reparameterization of Ψ (Pinheiro and Bates 1996; Pourahmadi 1999). Equality and inequality constraints can be accommodated by the Lagrangian rule provided by Theorem 6.1.1 in Clarke (1990, p. 228).

Using the subgradient approach of Rockafellar (1970), Koenker (2005) gives optimality conditions of the quantile regression problem for independent data. Here, we consider the gradient search algorithm for Laplace likelihood originally developed by Bottai and Orsini (2012) for censored Laplace regression (Bottai and Zhang 2010), which works as follows.

From a current parameter value the algorithm searches the positive semi-line in the direction of the gradient for a new parameter value at which the likelihood is larger. The algorithm stops when the change in the likelihood is less than a specified tolerance. Convergence is guaranteed by the continuity and concavity of the log-likelihood (8). Let us denote the gradient of $-\ell_{\text{app}}\{\theta, \sigma\}$ in (9) with $s(\theta^{k_1}, \sigma^0)$. The minimization steps are:

1. Set $\theta = \theta^0$; $\delta = \delta^0$; $\sigma = \sigma^0$; $k_1 = 0$.
2. If $\ell_{\text{app}}\{\theta^{k_1} - \delta^{k_1} s(\theta^{k_1}), \sigma^0\} \geq \ell_{\text{app}}\{\theta^{k_1}, \sigma^0\}$
 - (a) then set $\delta^{k_1+1} = a\delta^{k_1}$;
 - (b) else if $\ell_{\text{app}}\{\theta^{k_1}, \sigma^0\} - \ell_{\text{app}}\{\theta^{k_1} - \delta^{k_1} s(\theta^{k_1}), \sigma^0\} < \omega_1$
 - (i) then return θ^{k_1+1} ; stop;
 - (ii) else set $\theta^{k_1+1} = \theta^{k_1} - \delta^{k_1} s(\theta^{k_1})$; $\delta^{k_1+1} = b\delta^{k_1}$.
3. Set $k_1 = k_1 + 1$; go to step 2.

The algorithm requires setting the starting value of the parameter θ^0 and σ^0 , the initial step $\delta^0 > 0$, the contraction step factor $a \in (0, 1)$, the expansion step factor $b \geq 1$, and the tolerance $\omega_1 > 0$ for the change in ℓ_{app} . Additionally,

a check on the convergence of the parameter θ can be introduced in step (b) by verifying $\max |\delta^{k_1} s(\theta^{k_1})| < \omega_2$, where $\omega_2 > 0$ controls the tolerance. Least squares estimates can be used to initialize θ_x . Once the algorithm finds a solution for θ , the scale parameter σ can be estimated residually to obtain σ^1 or a second iterative loop as above can be initialized by setting $\sigma = \sigma^1$ and so forth until the change in the parameter is sufficiently small. Simulation results suggest that in most cases the number of iterations of the second loop is small (typically, less than four), as is the change in the estimate.

3.3 Interpretation of parameters

Consider the random intercept model introduced in (1). Its marginal distribution is given by $y \sim N(\mu, \psi_u^2 Z Z' + \psi^2 I)$. The parameters of such model have a straightforward interpretation: μ is the mean effect at the population level, ψ_u^2 is a measure of the dispersion of the cluster-specific random effects and proportional to the ICC, and ψ^2 is the error's variance.

Each cluster has a conditional distribution $N(\mu + u_i, \psi^2 I)$, therefore the 'atomic' τ th quantile is $q_{i.}(\tau) \equiv \mu + u_i + \psi \Phi^{-1}(\tau)$ for all $j = 1, \dots, n_i$, where Φ^{-1} denotes the inverse of the cumulative distribution function of a standard normal. Conditionally on u_i , the y_{ij} 's are independent. Thus, the τ th sample quantile $\hat{q}_{i.}(\tau)$ is an estimator of $q_{i.}(\tau)$, which, for large n_i ,

$$\hat{q}_{i.}(\tau) \sim N\left(q_{i.}(\tau), \frac{\tau(1-\tau)}{n_i [p_{y_{i.}|u_i}(q_{i.}(\tau))]^2}\right).$$

Note that this result is valid for any $0 < \tau < 1$ and continuous p . If we took the average of M such estimators we would obtain, when M is large,

$$\frac{1}{M} \sum_{i=1}^M \hat{q}_{i.}(\tau) \sim N(\bar{q}(\tau), v)$$

where $\bar{q}(\tau) = \frac{1}{M} \sum_i q_{i.}(\tau) \approx \mu + \psi \Phi^{-1}(\tau)$ and

$$\begin{aligned} v &= \frac{1}{M^2} \sum_i \frac{\tau(1-\tau)}{n_i [p_{y_{i.}|u_i}(q_{i.}(\tau))]^2} \\ &= \frac{1}{M^2} \sum_i \frac{\tau(1-\tau) 2\pi \psi^2}{n_i} \exp\{[\Phi^{-1}(\tau)]^2\}. \end{aligned}$$

It is intended that, under the above random intercept model, the approximation of the mean is valid on average (that is, $E_u\{\bar{q}(\tau)\}$). Thus, the fixed effects can be interpreted as a weighted average approximation of the population quantiles they target.

Let us now turn to the linear quantile mixed models. Our starting point is conditional. The marginal likelihood (4), estimated under (5) or (6), implicitly assumes that the τ th regression quantiles of the clusters 'gravitate' around a common regression quantile which, in the case of the normal intercept model, would be $\mu + \psi \Phi^{-1}(\tau)$, clearly different from the τ th quantile of the marginal model $\mu + \sqrt{\psi_u^2 + \psi^2} \Phi^{-1}(\tau)$. Oberhofer and Haupt's (2005) showed that the unconditional quantile estimator for dependent random variables is asymptotically unbiased.

The scale parameter σ does not have, in general, a straightforward interpretation since the use of the Laplace distribution for the conditional response follows from the convenience of manipulating a likelihood rather than from the observation that the data is indeed Laplacian. But what if it is? Consider the linear median mixed model ($\tau = 0.5$). The variance of the quantile estimator of $q_{i.}(0.5)$ based on normal approximations is $\frac{\tau(1-\tau)}{n_i [\frac{\tau(1-\tau)}{\sigma}]^2} = \frac{4\sigma^2}{n_i}$. The relative efficiency of this estimator under Laplacian and normal hypotheses is then $\frac{4\sigma^2}{n_i} / \frac{\pi\psi^2}{2n_i} = 8\sigma^2/\pi\psi^2$. Note that if $y \sim \text{AL}(\mu, \sigma, \tau)$ then $\text{var}(y) = \frac{\sigma^2(1-2\tau+2\tau^2)}{(1-\tau)^2\tau^2}$ (Yu and Zhang 2005), hence $\text{var}(y) = 8\sigma^2$ for $\tau = 0.5$. The median estimator would achieve the same asymptotic efficiency under the two distributions if, in terms of variances, $8\sigma^2/\psi^2 = \pi$ or, equivalently in terms of scale parameters, if $\sigma/\psi \approx 0.62$.

3.4 Inference

The LQMM in (2) as well as GB's random-intercept model assume constant scale parameter and, under this assumption, the asymptotics in heteroscedastic data would lead to incorrect inferential approximations. However, GB's standard error calculation relied on bootstrap resampling which makes large- n approximations irrelevant (Kim and Yang 2011). The inclusion of variance weights might represent a possible development to address such limitation in our models. See Koenker (2005, p. 94) for a discussion on likelihood ratio tests based on the asymmetric Laplace residuals.

Bootstrap represents a very flexible approach and is often used in quantile estimation (Parzen et al. 1994; Buchinsky 1995; He and Hu 2002; Kocherginsky et al. 2005; Bose and Chatterjee 2003; Koenker 2005; Feng et al. 2011; Canay 2011). Here, we consider a block bootstrap approach to assess the uncertainty in conditional quantile estimates. R bootstrap samples are obtained by resampling the index $i = 1, \dots, M$ with replacement. The standard errors are computed as the square root of the diagonal elements of the covariance matrix $\hat{V} = \text{Cov}(B)$, where B is a matrix $R \times (p+1)$ with row vectors of bootstrap estimates of $(\theta_x, \sigma)'$. Similarly, standard errors for Ψ can be obtained after opportunely transforming the bootstrap estimates of θ_z .

Model selection can be conducted by using traditional measures such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). A Laplacian likelihood ratio test can be used for testing linear hypotheses of the type $H_0: \theta_2 = 0$, with partitioned coefficients vector $\theta_x = (\theta'_1, \theta'_2)'$. See Koenker and Machado (1999) for additional details.

As mentioned previously, random-effects models allow conditional inferences in addition to marginal inferences (Robinson 1991; Lee and Nelder 2004). Prediction of the random effects vector u needs some elaboration. Here, we briefly illustrate a possible approach based on best prediction. Consider the simple case where $y = u + \varepsilon$. The best predictor of u is defined as the vector \tilde{u} that solves

$$\min_{\tilde{u}} E\{(\tilde{u} - u)^2\}. \quad (10)$$

The solution to (10) is the conditional expectation $E(u|y)$ (Ruppert et al. 2003). If we restrict the family of predictors to be linear and we consider the general LQMM $y = X\theta_x^{(\tau)} + Zu + \varepsilon^{(\tau)}$, the solution is the best linear predictor (BLP) of Zu

$$Z\tilde{u} = E(Zu) + \{\text{Cov}(Zu, y)\} \Sigma^{-1} \{y - E(y)\}. \quad (11)$$

The first term of expression (11) is zero and, following from the assumption that $u \perp \varepsilon^{(\tau)}$, $\text{Cov}(Zu, y) = Z\Psi^{(\tau)}Z'$. The last two terms of the product are the inverse of the matrix $\Sigma \equiv \text{Cov}(y) = Z\Psi^{(\tau)}Z' + \text{Cov}(\varepsilon^{(\tau)})$ and the vector $y - E(y) = y - X\theta_x^{(\tau)} - E(\varepsilon^{(\tau)})$. Expressions for mean and variance of the asymmetric Laplace are given in Yu and Zhang (2005). If in (11) we plug an estimate of θ and σ , we refer to (11) as estimated BLP (eBLP) and we write

$$\hat{u}_{\text{eBLP}}^{(\tau)} = \hat{\Psi}^{(\tau)}Z'\hat{\Sigma}^{-1}\{y - X\hat{\theta}_x^{(\tau)} - \hat{E}(\varepsilon^{(\tau)})\}. \quad (12)$$

For the sake of simplicity, consider the LMM $y = u + \varepsilon$, $u \sim N(0, I)$, $\varepsilon \sim N(0, I)$ which entails an ICC of 0.5. As seen previously, the random effects play a pure location-shift role. Therefore, we expect that in this case a good predictor for u in LQMMs provided a prediction \hat{u} that is close to the best linear unbiased prediction (BLUP) of u from a LMM.

In a small-scale simulation study, we generated $n = 10$ repeated observations from a LMM with $M = 50$ groups and we calculated the eBLP $\hat{u}_{\text{eBLP}}^{(\tau)}$ given in (12) for $\tau \in \mathcal{T}$, $\mathcal{T} = \{0.01, 0.02, 0.05, 0.1, \dots, 0.95, 0.98, 0.99\}$, and the estimated BLUP (eBLUP) \hat{u}_{eBLUP} for 200 replications. We then estimated the coefficients of the linear regression

$$E(\hat{u}_{\text{eBLP}}^{(\tau)} | \hat{u}_{\text{eBLUP}}) = a + b \cdot \hat{u}_{\text{eBLUP}}, \quad \tau \in \mathcal{T}.$$

The mean squared error for a and b was, respectively, 0.056 and 0.018. This was calculated as the squared difference between the estimated regression coefficients and those from a

perfect linear relationship, averaged over the simulation realizations and the number of τ 's. Selected results are presented in Fig. 1. In general, the median of the intercepts estimated from the model above was close to 0 at all estimated quantiles. The median of the slopes b was close to 1 for τ 's in the region of 0.5 but decreased to around 0.8 for $\tau = 0.01$ and $\tau = 0.99$. The spread of the slopes was larger for tail quantiles. These results remained unchanged when $M = 100$ and $n = 20$. However, to higher values of the intraclass correlation corresponded slopes closer to 1 at all estimated quantiles. Conversely, lower ICC values led to weaker association between $\hat{u}_{\text{eBLP}}^{(\tau)}$ and \hat{u}_{eBLUP} .

The residuals $\hat{e} = y - X\hat{\theta}_x^{(\tau)} - Z\hat{u}_{\text{eBLP}}^{(\tau)}$ did not satisfy the optimality condition of Koenker and Bassett (1978, Theorem 3.4). An *ad-hoc* adjustment might consist in constraining the estimation of u in (11) to satisfy such condition. The development of a predictor for the random effects that take into accounts these preliminary results is needed.

4 Simulation study

We carried out an extensive simulation study. Data were generated according to

$$y_{ij} = (\beta_0 + u_i) + (\beta_1 + v_i)x_{ij} + \beta_2 z_{ij} + (1 + \gamma x_{ij})e_{ij},$$

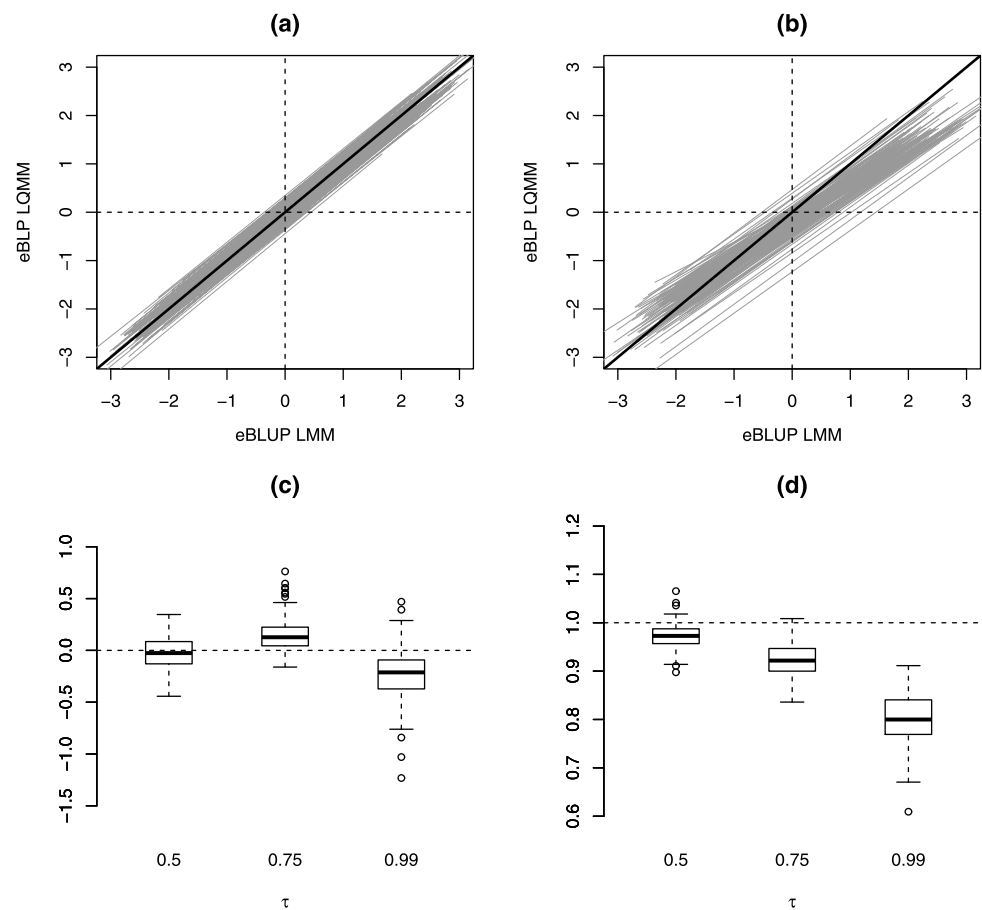
where $\beta = (100, 2, 1)'$, u_i and v_i are cluster-specific random effects, $x_{ij} = \delta_i + \zeta_{ij}$, $\delta_i \sim N(0, 1)$, $\zeta_{ij} \sim N(0, 1)$ and $z_{ij} \sim \text{Binom}(1, 0.5)$. To ensure a positive scale parameter and thus avoid quantile-crossing, δ_i and ζ_{ij} were generated as standard uniforms in all heteroscedastic models ($\gamma \neq 0$). A summary of the simulation scenarios for different data-generating distributions and sample sizes is given in Table 1.

Quantile mixed models of the form

$$y_i = X_i\theta_x^{(\tau)} + Z_i u_i + \varepsilon_i^{(\tau)},$$

where X_i is a $n \times 3$ matrix that collects a column of ones and the variables x and z , were estimated for $\tau \in \{0.5, 0.75, 0.9\}$. The matrix Z_i and the dimension of the random-effects vector u_i were defined according to each scenario, and so was the LQMM covariance matrix $\Psi^{(\tau)}$ (i.e., either an identity or a symmetric matrix). A Gauss–Hermite quadrature with $K = 11$ nodes was used to approximate the marginal log-likelihood as in (5). For data generated under models (14–16), the number of nodes was set to $K = 17$. All models were estimated with the gradient search algorithm described in Sect. 3.2, with the exception of those applied to data generated under scenarios (22–23), which were estimated using a Nelder–Mead algorithm. The latter was also used for models (1–6) to compare derivative-free with gradient-based optimization. The tolerance parameter and the maximum number of iterations were set to, respectively, 10^{-3} and 500 for

Fig. 1 Plot of best linear and best linear unbiased predictions (grey lines) of u for (a) $\tau = 0.5$ and (b) $\tau = 0.99$; and boxplot of the intercepts (c) and slopes (d) from the regression model $E(\hat{u}_{\text{eBLP}}^{(\tau)} | \hat{u}_{\text{eBLP}}) = a + b \cdot \hat{u}_{\text{eBLP}}$ for three quantiles in 200 replicated data sets



the likelihood, and to 10^{-5} and 10 for the scale parameter σ . The standard deviation of the response was used as initial step size. The contraction and expansion step factors were set to 0.5 and 1 respectively. Starting values for θ_x were obtained via least squares estimation. For models fitted to contaminated data (22–23), Gauss–Laguerre quadrature was also assessed. In addition, tail quantiles $\tau \in \{0.01, 0.05, 0.95, 0.99\}$ were estimated for data generated under scenarios (1, 3, 5) using a sample size $n = 10$ and $M = 100$, and a likelihood tolerance parameter equal to $5 \cdot 10^{-3}$.

We also assessed the coverage rate of the median estimator's confidence intervals at the nominal 90 % level. Data were generated as in scenario (3) (Table 1) but with fixed design points (i.e., drawn only once), either for $\gamma = 0$ or $\gamma = 0.25$ and for three sample sizes (n, M) : (5, 50), (10, 100), and (10, 200). The standard errors were calculated using $R = 50$ bootstrap replications. The estimates were obtained using a Gauss–Hermite quadrature with $K = 17$ nodes and Nelder–Mead optimization.

Finally, we evaluated the Gauss–Hermite quadrature for several quadrature grid sizes. Data were generated from $y_{ij} = (\beta_k + u_{i,k})x_{ij,k} + e_{ij}$, $u_{i,k} \sim N(0, 5)$, $e_{ij} \sim N(0, 5)$ with $n = 5$ and $M = 50$. The number of fixed and random

effects k ranged from 1 (intercept model) to 6, with $x_{ij,k} = \delta_i + \zeta_{ij}$ as described above for $k = 2, \dots, 6$. LQMMs were estimated for $\tau \in \{0.5, 0.75, 0.9\}$ using $K = 5$ or $K = 17$ nodes.

In all cases, datasets were generated independently 500 times. In what follows, we report selected results for the sake of brevity. All results are available upon request from the first author.

For each scenario in Table 1, we assessed the relative bias, averaged over the simulation realizations, and reported the results in Table 2. Overall, the performance of the LQMM estimator was satisfactory, except in specific cases. Estimation bias was low at all considered quantiles for one- or two-dimensional random-effects models. However, an increase of bias in the slope of x was observed in data generated with correlated random effects, in particular when using t_3 distributions (scenario 19¹). The bias for this coefficient became severe when u and v were drawn from skewed, in addition to heavy-tailed, distributions (scenarios 20 and 21), which, of course, was due to the use of symmetric quadra-

¹For this scenario, we increased the number of quadrature nodes K from 11 to 17. The relative bias decreased from 0.135 to 0.051 ($\tau = 0.75$) and from 0.459 to 0.075 ($\tau = 0.9$).

Table 1 Simulation study scenarios. Distributions: $N(0, \psi^2)$, normal with mean 0 and variance ψ^2 ; t_g , t with g degrees of freedom; χ_g^2 , chi-square with g degrees of freedom; St_g , skew t with g degrees of freedom (Azzalini and Capitanio 2003)

Model description	(n, M)	u	v	e	γ
(1) location-shift symmetric	(5, 50)	$N(0, 5)$	—	$N(0, 5)$	0
(2) location-shift symmetric	(5, 50)	t_3	—	$N(0, 5)$	0
(3) location-shift heavy-tailed	(5, 50)	$N(0, 5)$	—	t_3	0
(4) location-shift heavy-tailed	(5, 50)	t_3	—	t_3	0
(5) location-shift asymmetric	(5, 50)	$N(0, 5)$	—	χ_2^2	0
(6) location-shift asymmetric	(5, 50)	t_3	—	χ_2^2	0
(7) location-shift symmetric $\text{cor}(u, v) = 0$	(5, 50)	$N(0, 5)$	$N(0, 5)$	$N(0, 5)$	0
(8) location-shift heavy-tailed $\text{cor}(u, v) = 0$	(5, 50)	t_3	t_3	t_3	0
(9) location-shift symmetric	(10, 100)	$N(0, 5)$	—	$N(0, 5)$	0
(10) location-shift symmetric	(20, 200)	$N(0, 5)$	—	$N(0, 5)$	0
(11) location-shift heavy-tailed $\text{cor}(u, v) = 0$	(10, 100)	t_3	t_3	t_3	0
(12) location-shift heavy-tailed $\text{cor}(u, v) = 0$	(20, 200)	t_3	t_3	t_3	0
(13) location-shift asymmetric $\text{cor}(u, v) = 0$	(20, 200)	t_3	t_3	χ_2^2	0
(14) heteroscedastic symmetric	(10, 100)	$N(0, 5)$	—	$N(0, 5)$	0.25
(15) heteroscedastic heavy-tailed	(10, 100)	t_3	—	t_3	0.25
(16) heteroscedastic asymmetric	(10, 100)	t_3	—	χ_2^2	0.25
(17) location-shift symmetric $\text{cor}(u, v) > 0$	(10, 100)	$N(0, 5)$	$N(0, 5)$	$N(0, 5)$	0
(18) location-shift symmetric $\text{cor}(u, v) < 0$	(10, 100)	$N(0, 5)$	$N(0, 5)$	$N(0, 5)$	0
(19) location-shift heavy-tailed $\text{cor}(u, v) > 0$	(10, 100)	t_3	t_3	t_3	0
(20) location-shift heavy-tailed $\text{cor}(u, v) < 0$	(10, 100)	St_3	St_3	t_3	0
(21) location-shift heavy-tailed $\text{cor}(u, v) > 0$	(10, 100)	St_3	St_3	t_3	0
(22) location-shift with 5 % contamination	(10, 100)	$N(0, 5)$	—	$\chi_2^2 + N(0, 50)$	0
(23) location-shift with 5 % contamination	(10, 100)	$N(0, 5) + N(0, 50)$	—	χ_2^2	0

ture weights. Mean CPU time², averaged over the simulation realizations and the number of τ 's, ranged from 0.02 seconds in simpler scenarios to 2.80 seconds in scenarios with more complex random-effects structures or larger sample sizes.

Outlier contamination, either in the random intercept or in the error term (scenarios 22 and 23), did not impact the estimation bias. However, for data generated under scenario (23), the Monte Carlo variance of the estimated intercepts using Gauss–Laguerre quadrature was 51 % ($\tau = 0.5$), 39 % ($\tau = 0.75$), and 28 % ($\tau = 0.9$) smaller than that calculated when using normal quadrature weights (results not shown).

Estimation bias in scenarios (1–6) remained unchanged when using derivative-free estimation (results not shown).

As expected, tail quantiles had higher estimation bias, in particular for $\tau = 0.01$ and $\tau = 0.99$, though results (not shown) differed markedly depending on the coefficient being estimated. The relative absolute bias, averaged over τ

and the three scenarios, was 1 % for the intercept and 0.6 % for x 's slope but 24 % for the binary covariate's coefficient.

Coverage rate and average length of bootstrap confidence intervals are reported in Table 3. Coverage rates were close to the nominal level in most cases. As expected, the average length of the confidence intervals decreased with increasing sample size and was larger in the heteroscedastic scenario. In the Bayesian context, credible intervals for the location parameter of AL-distributed variates have been found to achieve nominal coverage at several quantiles (Lum and Gelfand 2012).

Relative absolute bias, total number of iterations and CPU time, averaged over the simulation realizations and the number of τ 's, are reported in Table 4. While the number of iterations grows linearly, computation time increases exponentially with increasing q . Bias is low at all grid sizes and, for a given dimension q , it remains approximately unchanged when increasing the number of nodes from 5 to 17. Based on these results, we recommend using a small number of quadrature nodes (say, between 5 and 11) especially when fitting models with more than three random effects. However, as noted with regard to scenario 19 (Table 2), ran-

²All computations were performed on a 64-bit operating system machine with 16 Gb of RAM and quad-core processor at 2.93 GHz.

Table 2 Relative bias of $\hat{\theta}_x^{(\tau)}$ for $\tau \in \{0.5, 0.75, 0.9\}$

Model	Intercept			x			z		
	0.5	0.75	0.9	0.5	0.75	0.9	0.5	0.75	0.9
(1)	−0.000	0.001	0.002	−0.002	−0.003	0.000	0.007	0.019	−0.018
(2)	0.000	0.001	0.000	0.001	−0.004	0.001	0.018	0.041	−0.017
(3)	0.000	0.003	0.008	0.005	0.004	−0.002	0.013	0.034	0.013
(4)	0.000	0.002	0.004	0.004	−0.000	−0.002	−0.006	0.034	−0.024
(5)	0.002	0.005	0.003	0.001	0.004	0.001	0.010	0.056	0.072
(6)	0.002	0.003	0.002	0.004	0.007	0.011	−0.018	0.055	0.048
(7)	−0.000	0.001	0.003	0.005	0.006	−0.009	−0.002	0.047	−0.107
(8)	−0.000	0.010	0.039	−0.013	−0.013	−0.004	0.038	−0.002	−0.038
(9)	−0.000	0.000	0.002	0.002	−0.001	−0.001	−0.007	0.002	−0.007
(10)	−0.000	−0.001	0.002	−0.000	0.002	−0.001	−0.002	0.001	−0.008
(11)	−0.000	0.011	0.057	−0.010	0.006	−0.006	−0.006	−0.008	0.008
(12)	0.000	0.009	0.021	−0.016	−0.018	−0.020	−0.005	−0.011	−0.005
(13)	0.002	0.006	0.015	0.012	−0.009	0.018	−0.004	−0.007	−0.009
(14)	−0.000	0.000	0.001	−0.006	−0.005	−0.072	0.011	0.012	0.001
(15)	0.000	0.001	0.002	0.009	0.010	−0.005	−0.011	−0.007	−0.018
(16)	0.002	0.001	0.000	0.018	0.034	0.024	−0.005	−0.002	−0.006
(17)	0.000	−0.000	0.003	0.002	−0.108	0.017	0.004	0.032	−0.006
(18)	−0.000	0.002	0.005	−0.013	−0.045	−0.089	−0.010	−0.000	−0.034
(19)	0.000	0.011	0.068	−0.009	0.135	0.459	−0.008	−0.024	−0.014
(20)	0.006	0.010	0.016	0.328	0.258	0.218	−0.003	−0.002	−0.018
(21)	0.012	0.015	0.019	0.624	0.566	0.614	0.004	0.013	0.004
(22)	0.003	0.005	0.010	0.001	0.003	0.005	−0.002	−0.009	−0.034
(22) ^a	0.003	0.005	0.009	0.001	0.005	0.004	−0.006	−0.010	−0.050
(23)	0.003	0.003	0.004	0.001	−0.000	−0.007	−0.002	−0.001	−0.006
(23) ^a	0.003	0.003	0.003	−0.001	−0.002	−0.005	−0.003	0.003	0.002

^aRobust (Gauss–Laguerre) random-effects estimation.

Table 3 Coverage rate and mean length of the confidence intervals based on 50 bootstrap replications for the median estimator at nominal 90 % level

(n, M)	Location-shift ($\gamma = 0$)			Heteroscedastic ($\gamma = 0.25$)		
	(5, 50)	(10, 100)	(10, 200)	(5, 50)	(10, 100)	(10, 200)
Coverage (%)						
Intercept	91.6	91.2	94.4	91.0	92.0	92.4
x	93.0	93.2	94.2	90.8	91.0	94.2
z	90.6	91.0	94.4	93.6	94.6	90.2
Length						
Intercept	1.7	1.4	1.2	2.2	1.6	1.2
x	0.4	0.2	0.2	1.5	0.8	0.6
z	0.8	0.4	0.3	0.9	0.4	0.3

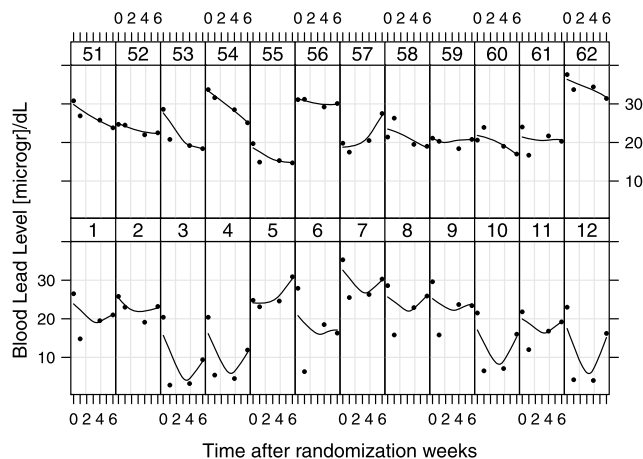
dom effects distributions characterized by heavy tails may require using a larger number of nodes to reduce the estimation bias.

Substantial reductions in the size of the quadrature grid for a given level of accuracy can be obtained by employing efficient quadrature rules, such as, for example, integra-

tion on sparse grids (Heiss and Winschel 2008) and nested integration rule for Gaussian weights (Genz and Keister 1996). Also, adaptive rescaling (Pinheiro and Bates 1995; Pinheiro and Chao 2006) may improve the estimation when the variance of the random effects is large. These are topics of current research by the first author.

Table 4 Relative absolute bias of $\hat{\theta}_x^{(\tau)}$ for $\tau \in \{0.5, 0.75, 0.9\}$, number of iterations and time to convergence

Nodes K	Dimension q	Grid size	Relative absolute bias	Iterations	CPU time (seconds)
5	1	5×1	0.000	37	0.01
	2	25×2	0.006	50	0.05
	3	125×3	0.020	56	0.31
	4	625×4	0.014	61	2.36
	5	3125×5	0.014	65	18.10
	6	15625×6	0.026	71	126.83
17	1	17×1	0.000	39	0.02
	2	289×2	0.004	52	0.46
	3	4913×3	0.005	58	11.77
	4	83521×4	0.023	65	332.02

**Fig. 2** Trellis plot of blood lead level in 24 children treated with succimer (subjects 1–12) or with placebo (subjects 51–62)

5 Examples

5.1 Treatment of lead-exposed children

In this section, we revisit a placebo-controlled, double-blind, randomized trial of succimer (a chelating agent) in children with blood lead levels (BLL) of 20–44 $\mu\text{g}/\text{dL}$ conducted by the Treatment of Lead-Exposed Children (TLC) Trial Group (2000). Lead exposure is known to increase the risk of encephalopathy and, in general, to impair cognitive function. The trial was designed to test the hypothesis that children with moderate BLL who were treated with succimer would score higher than children given placebo on a number of behavioral and cognitive tests. Drug treatment was later found not to improve psychological test scores (Rogan et al. 2001).

The original data consist of repeated measurements of BLL observations on an initial sample of 780 children (396 in the succimer group and 384 in the placebo group). Here, we analyze a subset of the data for $M = 100$ children and for a shorter time period. Our results, therefore, may not

be extended to draw general conclusions on the original study. The dataset, available at <http://biosun1.harvard.edu/~fitzmaur/ala/tlc.txt>, provides BLL obtained at baseline (or week 0), week 1, week 4, and week 6 on children randomly assigned to treatment (50) or placebo (50) group.

A Trellis plot (Sarkar 2008) of selected individual blood lead trajectories is shown in Fig. 2. The initial drop in BLL is followed by a rebound after 1 week of treatment in the succimer group (all sample curves are reported in Treatment of Lead-Exposed Children (TLC) Trial Group 2000). A monotonic pattern, in contrast, describes observed BLL in the placebo group. The overall distribution of BLL in each week, conditional on treatment status, is presented in Fig. 3. Non-normal features of the conditional distribution of BLL at each time point and in both groups are apparent.

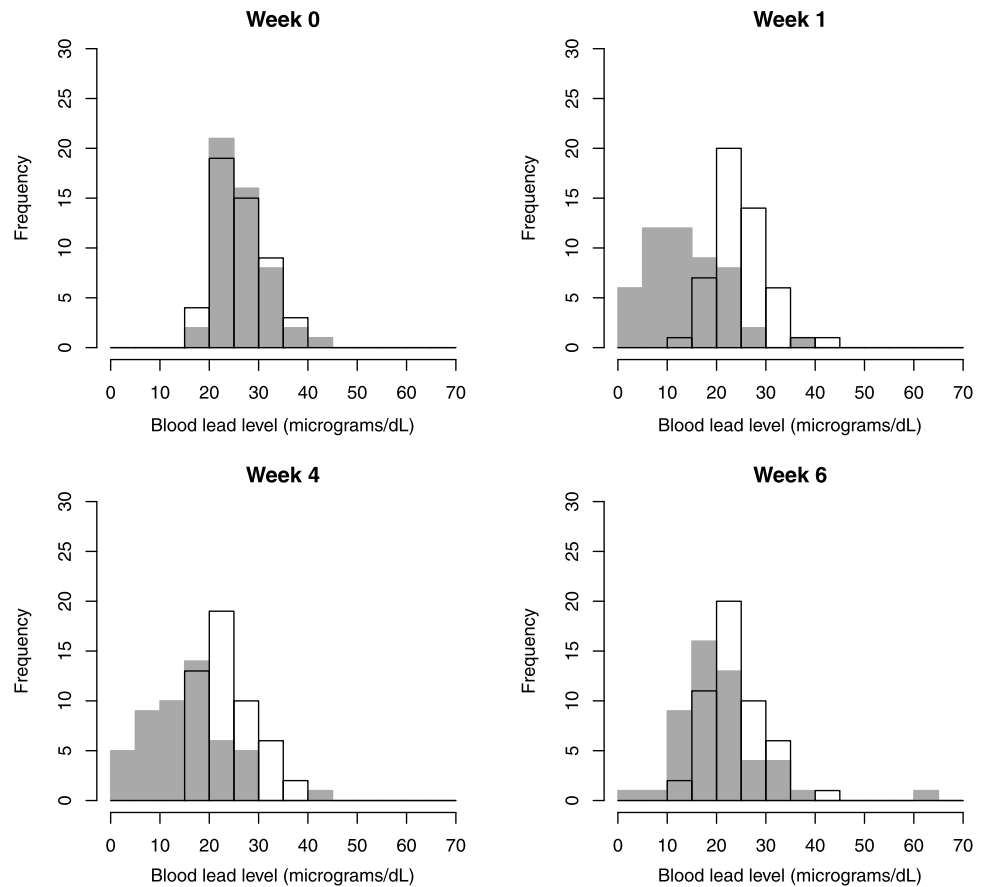
We modeled the temporal trajectories by using natural cubic splines with three degrees of freedom and break-points placed at 0, 4 and 6 weeks. The B-spline basis matrix $T = [T_1|T_2]$ was then included in the quantile mixed model

$$y_i = X_i \theta_x^{(\tau)} + Z_i u_i + \varepsilon_i^{(\tau)},$$

where X_i is a 4×6 matrix that collects a column of ones, a dummy variable for treatment (baseline: placebo), the basis splines T_1 and T_2 and their interaction with treatment. The matrix Z_i includes the same variables as X_i except for those associated with T_2 (the inclusion of a random slope for T_2 was not found to significantly improve the fit of the model). Therefore, θ_x is a 6×1 vector of fixed coefficients and u_i is a 4×1 vector of subject-specific random effects.

We assumed a general positive-definite matrix and approximated the log-likelihood using a Gauss–Hermite quadrature as in (5) with $K = 7$ nodes, giving a quadrature grid of size 2401×4 , for 5 quantiles $\tau \in \mathcal{T}$, $\mathcal{T} = \{0.1, 0.25, 0.5, 0.75, 0.9\}$. The objective function was optimized by using the gradient search algorithm described in Sect. 3.2. Standard errors were computed using $R = 30$ bootstrap replications.

Fig. 3 Histograms of blood lead level in 100 children treated with succimer (*grey*) or with placebo (*white*) at 0, 1, 4, and 6 weeks from randomization



The estimated regression quantiles and their 95 % confidence intervals (error bars) are shown in Fig. 4. The least squares solution from a LMM (dashed line) and the 95 % confidence interval (dotted lines) are provided for comparison.

The intercepts can be interpreted as the quantiles τ , $\tau \in \mathcal{T}$, of the distribution of BLL in untreated children at the beginning of the study. The estimated treatment effect shows that the BLL distributions in the two groups at baseline differ markedly at lower quantiles and moderately at the centre.

The estimated slopes for spline T_1 in the placebo group show a fall of BLL over the 6-week time period at approximately uniform rates across the distribution, with levels comparable to the average rate. The negative estimates of T_2 's coefficients indicate a slightly steeper decrease of BLL after 4 weeks at rates that are, again, similar across the distribution.

In the treatment group, there is a generalized fall of BLL levels at all quantiles but at faster rates than those in the placebo group. The positive coefficients for spline T_2 reflect the rebound that we observed in Fig. 2. The magnitude of this effect is weaker at higher levels of lead concentration.

The analysis of the so-called Lehman-Doksum quantile treatment effect (Doksum 1974; Lehmann 1975; Koenker

and Xiao 2002) has particular relevance in placebo-controlled studies as the one considered here. In the case of independent observations, formal testing of location-shift and location-scale-shift hypotheses can be carried out, for example, by using Khmaladze-type tests (Koenker and Xiao 2002; Koenker 2005). To the best of our knowledge, these are not available for correlated data though it has been suggested that an extension of these tests to the analysis of cluster-specific distributions may be contemplated (Koenker 2005, p. 281).

We conclude this example with a brief analysis of the effects related to the random part of the model. Variance components and correlations are shown in Table 5. Mean and median show similar levels of variability and covariability of the random effects. There is an increasing trend in the variance of the intercepts from lower to higher quantiles as well as for the effects associated with T_1 's slopes in both groups. In contrast, intercepts in the treatment group show decreasing variance from lower to higher quantiles. In general, the correlation between random effects is substantial at all quantiles and between most of the random effects. Box-centile plots of the random effects predicted with the eBLP in (12) are shown in Fig. 5.

Fig. 4 Regression quantiles for blood lead level and 95 % confidence intervals (error bars). Least squares estimates (*dashed lines*) and 95 % confidence intervals (*dotted lines*) are reported

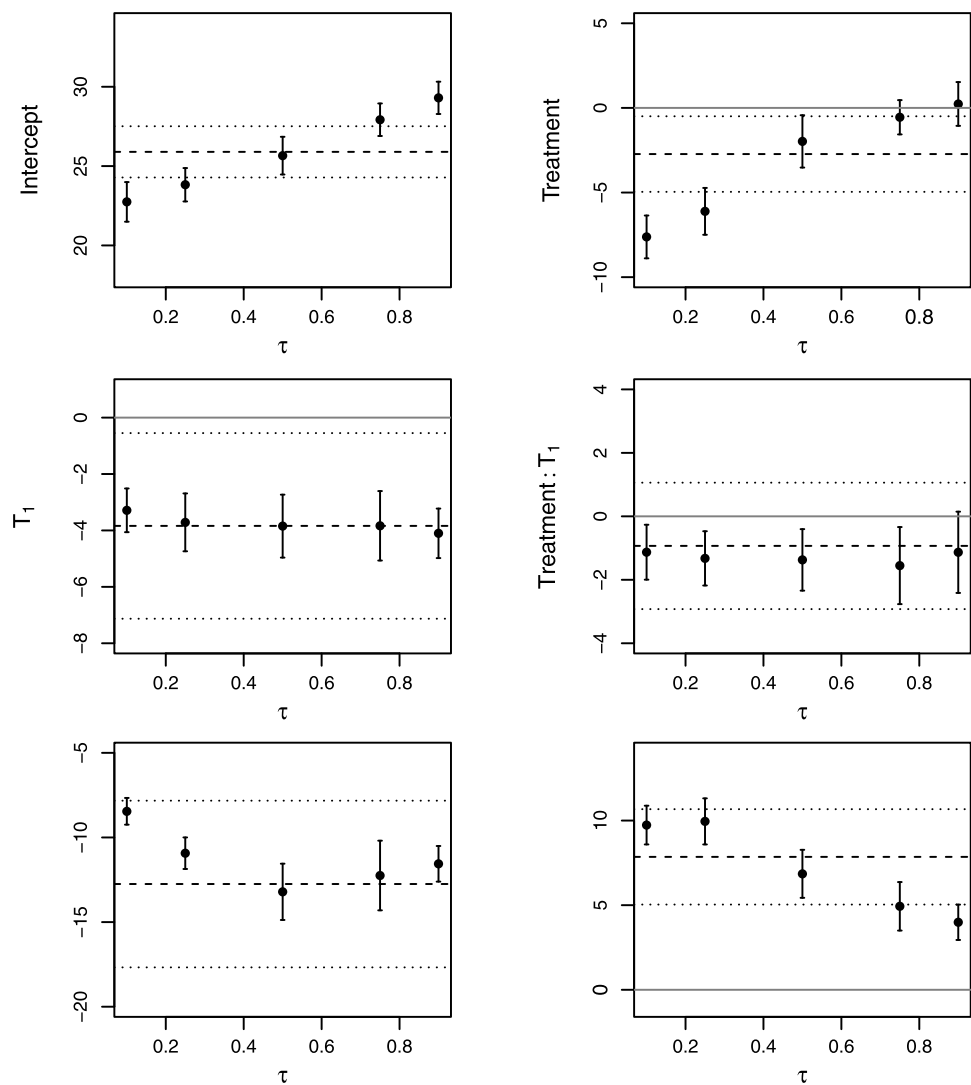
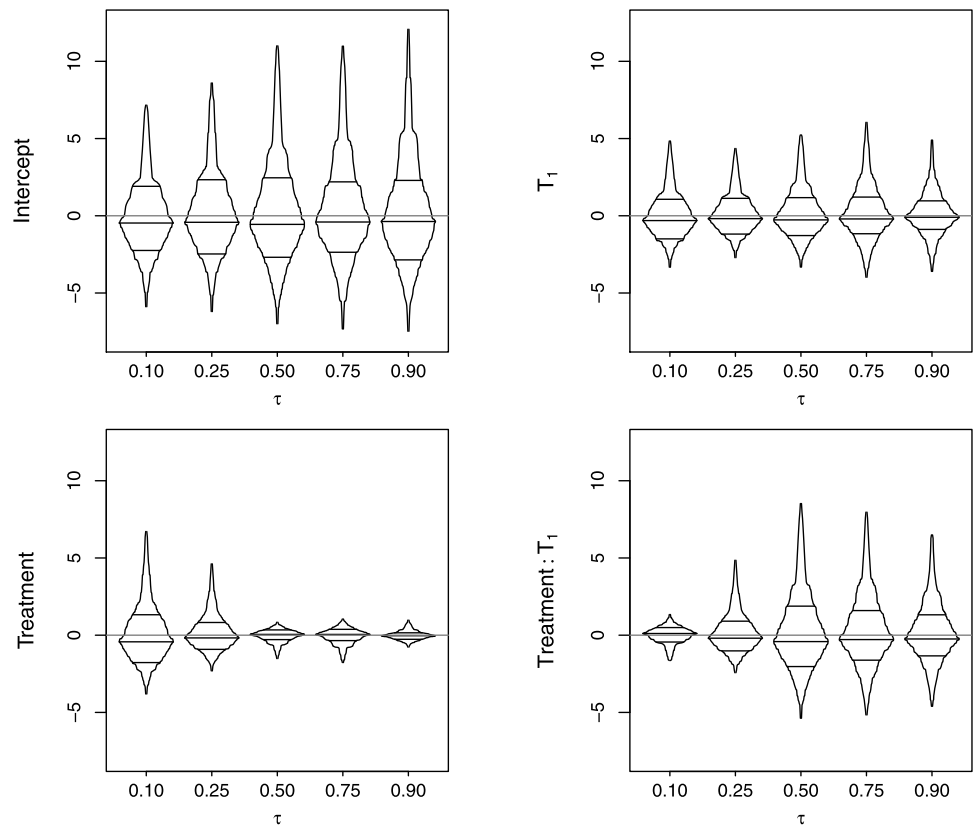


Table 5 Correlation matrix of the random effects for intercept (Int), basis spline T_1 , treatment variable (Trt) and their interaction term (Trt: T_1). Variances are reported in brackets

	Int	T_1	Trt	Trt: T_1	Int	T_1	Trt	Trt: T_1
	$\tau = 0.1$				$\tau = 0.25$			
Int	(11.05)				(13.40)			
T_1	0.84	(4.51)			0.86	(3.19)		
Trt	0.62	0.89	(9.93)		0.48	0.84	(4.92)	
Trt: T_1	-0.64	-0.65	-0.52	(1.34)	0.55	0.87	0.98	(4.96)
	$\tau = 0.5$				Mean			
Int	(17.49)				(16.38)			
T_1	1.00	(4.03)			1.00	(2.03)		
Trt	-0.59	-0.61	(1.00)		-0.44	-0.44	(4.91)	
Trt: T_1	0.95	0.93	-0.49	(10.79)	0.74	0.74	0.28	(23.56)
	$\tau = 0.75$				$\tau = 0.9$			
Int	(19.48)				(31.54)			
T_1	0.72	(7.28)			0.40	(10.52)		
Trt	-0.64	-0.39	(1.30)		0.25	0.15	(1.95)	
Trt: T_1	0.76	0.89	-0.75	(13.31)	0.58	0.94	0.04	(15.27)

Fig. 5 Box-centile plots of predicted child-specific random effects for blood lead level



5.2 A-level Chemistry scores

The dataset consists of A-level Chemistry examination scores in 1997 from 31,022 students in 2410 English schools (Fielding et al. 2003). Additional variables included in the dataset were gender, age in months centered at 222 months (18.5 years), average General Certificate of Secondary Education (GCSE) score centered at mean (6.3) and school identifier.

The outcome is a discrete variable and bounded between 0 and 10. Therefore, the following analysis is to be considered mainly for illustration purposes. To model the quantiles of A-level Chemistry scores, we considered the random-intercept model

$$y_i = X_i \theta_x^{(\tau)} + u_i + \varepsilon_i^{(\tau)},$$

where X_i is a $n_i \times 4$ matrix, $i = 1, \dots, 2410$, that collects a column of ones, a variable for age (range -6 to 5), a dummy variable for sex (baseline: male), and a variable for mean GCSE score as prior attainment (range -6.3 to 1.7). A random intercept u_i was introduced to capture the heterogeneity between schools whose number of students n_i varied from 1 to 188 (median 8).

We approximated the log-likelihood using a Gauss–Hermite quadrature as in (5) with $K = 9$ nodes for 7 quantiles $\tau \in \mathcal{T}$, $\mathcal{T} = \{0.1, 0.25, 0.4, 0.5, 0.6, 0.75, 0.9\}$ and optimized the objective function via Nelder–Mead. Standard

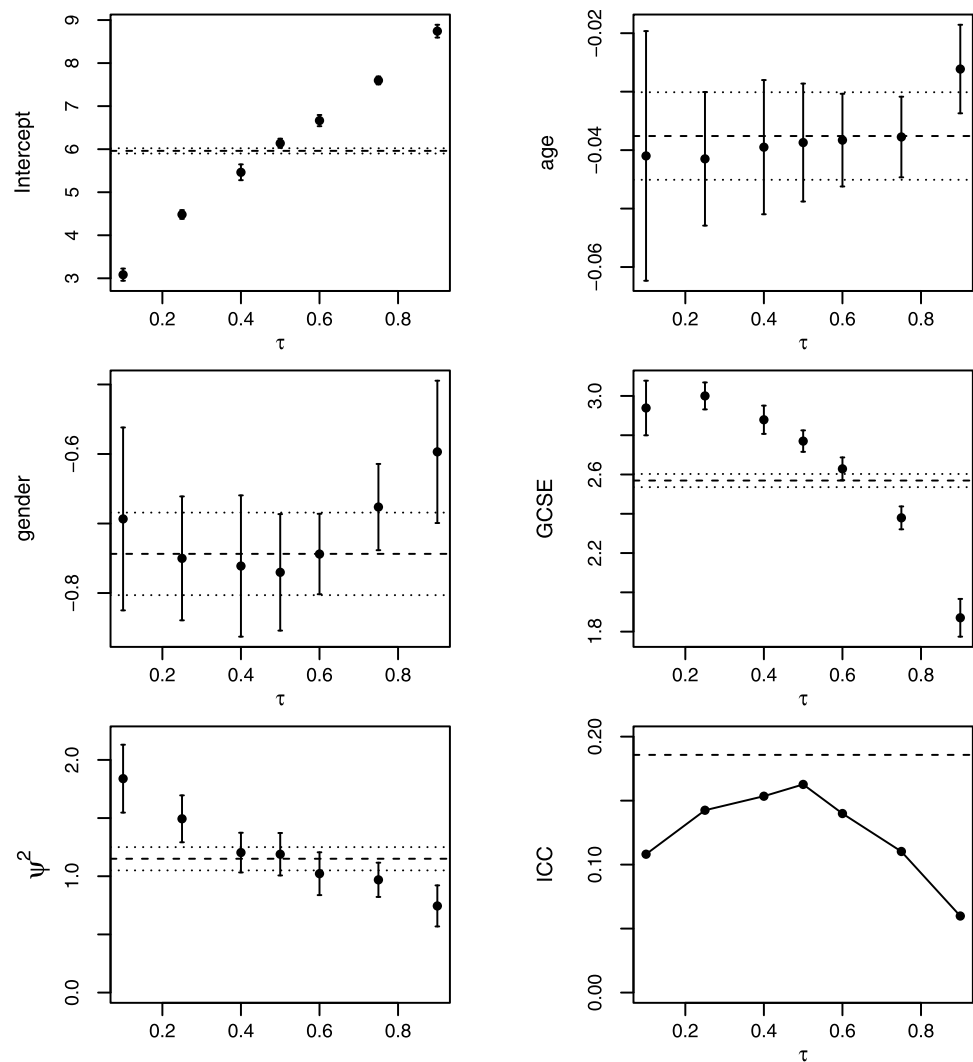
errors were computed using $R = 20$ bootstrap replications. In Fig. 6, the estimated regression quantiles and their 95 % confidence intervals (error bars) are depicted, together with the variance of the random intercepts, ψ^2 , and the ICC which are shown in the two plots at the bottom of the figure. The least squares solution from a linear mixed-effects model (dashed line) and the 95 % confidence interval (dotted lines) are provided for comparison.

The intercepts can be interpreted as the quantiles τ , $\tau \in \mathcal{T}$, of the distribution of Chemistry scores in 18.5 year-old males with a GCSE score equal to 6.3. The mean and median scores are very close and the monotonic increase of the quantiles denotes a uniform distribution.

The regression coefficients for age and sex are approximately constant and negative across quantiles and close to the corresponding mean age (-0.037) and sex (-0.74) effects. That is, older individuals and females perform less satisfactorily than younger peers and males, respectively, regardless of their rank in the Chemistry score distribution. However, at higher quantiles ($\tau = 0.9$) and therefore for high performing students, there is some indication that these effects may be weaker.

Prior achievement, as measured by mean GCSE results, is beneficial on average Chemistry score. For one point increase of GCSE mean score, A-level Chemistry score increases by 2.57 points. However, the positive effect is much stronger at lower quantiles (low performing students) than at

Fig. 6 Regression quantiles, variance ψ^2 and intraclass correlation for A-level Chemistry scores and 95 % confidence intervals (error bars). Least squares estimates (*dashed lines*) and 95 % confidence intervals (*dotted lines*) are reported



higher quantiles (high performing students), with decreasing magnitude inbetween.

Finally, the between-school variability shows a decreasing pattern from lower to higher quantiles. It is interesting to note that, in contrast, the ICC follows an inverted U curve. That is, the intra-school correlation is higher at the center of the outcome distribution where, proportionally, the within-school variability is smaller. Differences between schools, therefore, might play a less important role in explaining the total variability in below- and above-the-average students' performance. If these results were to be confirmed by other analyses, this would obviously have an important implication for education policies at the national level.

5.3 Doubly robust meta-analysis

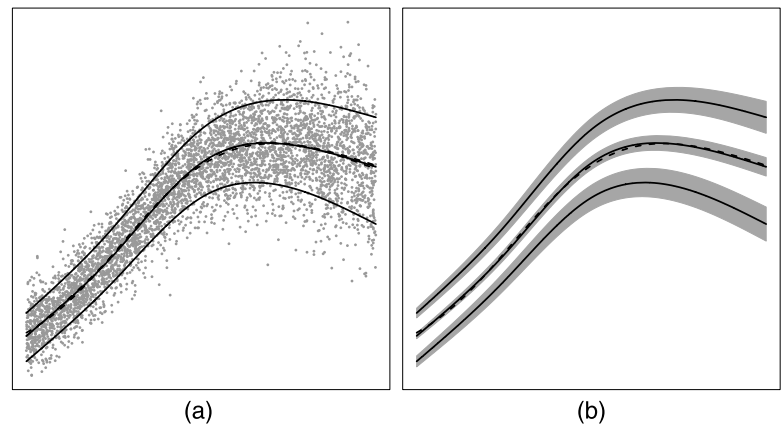
Let us consider the case in which M aggregate measures or summaries for a common underlying outcome are available but the raw data are not. These could be, for example, standard deviations coming from M experiments aimed at

assessing the precision of an instrument. Some instruments might have a much higher precision than others. We are interested in summarizing or pooling these studies by estimating one overall measure of the precision β_0 and estimating the deviation u_i , $i = 1, \dots, M$, from β_0 . Suppose each experiment has a valid sample of size n_i . Rather than assuming a normal random-effects model (DerSimonian and Laird 1986), we could consider estimating the random-effects median model

$$y_i = \beta_0 + u_i + \varepsilon_i^{(0.5)} \quad (13)$$

where y_i is the effect size for study i , $i = 1, \dots, M$ and u_i is a random effect that measures the deviation of the i th study effect from the population median β_0 . In addition, we could introduce Laplacian random effects in model (13). This would entail a double-robustness of the model: to unusual deviations from β_0 and to unusual deviations from $\beta_0 + u_i$. To take into account the different weight of the studies, the log-likelihood to be max-

Fig. 7 Smoothed regression quantiles ($\tau \in \{0.1, 0.5, .9\}$) with data points (a) and 95 % confidence bands (b). The true mean function is represented by a dashed line



imized would be $\sum_{i=1}^M \Pi(n_i) \ell_{\text{app},i}(\theta, \sigma)$, where $\Pi(\cdot)$ is some function of the size of the samples. See Demidenko (2004) for a normal-Laplace regression, where normality is assumed for ε . Model (13) would not have an advantage over existing meta-analysis models if the effect size has a sampling distribution approximately normal (e.g., as in the case of the sample mean).

5.4 Smoothing splines

Nonlinear QR for dependent data has received some attention in recent years. Karlsson (2008) developed a weighted approach to nonlinear QR estimation for repeated measures. Wang (2012) extended GB's model to account for nonlinear effects in normal-Laplace models.

Here, we show another possible application of the LQMMs presented in Sect. 2. We generated $n = 10$ observations clustered within $M = 500$ groups using the nonlinear heteroscedastic model

$$y_{ij} = \exp\left\{\frac{5x_{ij}}{1+2x^2}\right\} + u_i + x_{ij}\varepsilon_{ij}, \quad (14)$$

where $x_{ij} \sim U(0, 1)$, $u_i \sim N(0, 0.25)$ and $\varepsilon_{ij} \sim N(0, 1)$, independently. In this model, the within-group correlation $\text{cor}(y_{ij}, y_{ij'})$, $j \neq j'$, will vary in the range 0.2–1, depending on x_{ij} .

We approximated the nonlinear function in (14) with a natural cubic spline model with four degrees of freedom, break-points placed at the quartiles of x and boundary knots at the extremes of x 's range. The B-spline basis matrix S was then included in the quantile mixed model $y = S\beta^{(\tau)} + u + \varepsilon^{(\tau)}$. Fitted regression models estimated with a 9-knot Gauss–Hermite quadrature and 95 % confidence bands estimated by using 50 bootstrap replications are shown in Fig. 7 for three quantiles $\tau \in \{0.1, 0.5, 0.9\}$. Due to the symmetry of the error term, the estimated median and the (true) mean are very close. However, the spread of the distribution differs across the values of x .

5.5 Spatial modeling

Spatial modeling refers, loosely, to the case in which geographical information is introduced in the model. This includes the case in which x is a vector of geographical coordinates or, more simply, when the grouping factor is some geographical unit (e.g., postal codes, wards or counties). In the latter case, each random term u of the LQMM would be associated with a small-area effect and the resulting variance-covariance matrix would be interpreted as a spatial correlation matrix at the quantile of interest. The modeling of Ψ , therefore, should take into account the spatial association between areas (e.g., contiguity). See for example Lee and Neocleous (2010) for an application of Bayesian quantile regression (Yu and Moyeed 2001) to environmental epidemiology using the results of Machado and Santos Silva (2005). A Bayesian spatial quantile regression approach using the asymmetric Laplace is proposed also by Lum and Gelfand (2012).

Quantile smoothing of surfaces and related inference is a recent topic (He et al. 1998; He and Portnoy 2000; Koenker and Mizera 2004). An application of triogram smoothing splines to poverty mapping is described by Geraci and Salvati (2007). A Bayesian quantile modeling of ozone concentration surfaces is given by Reich et al. (2010b).

Suppose we want to estimate the quantiles of an outcome under the following nonparametric model:

$$y_i = f(x_i) + \varepsilon_i,$$

where x_i' is a vector of geographical coordinates. We could introduce a bivariate smoother for f as, for example, a generalized covariance function with quadratic penalty and adopt an approach as described in Ruppert et al. (2003) or a triogram smoothing spline model (Koenker and Mizera 2004) with L_1 penalized coefficients. Differently from the classical L_2 penalty used for estimating nonparametric mean functions, L_1 penalization (Koenker et al. 1994) has a more direct and natural application to quantiles from

a modeling and estimation standpoints. In both circumstances, however, the random effects would play the role of spline coefficients and the type of quadrature to be used would naturally follow the metric of the associated penalty term.

6 Final remarks

The growing interest in the theory and applications of regression methods for quantiles of outcome variables of research interest is to a large extent a consequence of the insight and ease of interpretation that only quantiles can provide. The present paper extends the work by Geraci and Bottai (2007) to the general linear quantile mixed effects regression model; discusses relevant aspects of estimation, testing, and computation in detail; illustrates the use of the linear quantile mixed effects regression in two real data examples; shows some of the results of an extensive simulation study; and indicates possible avenues of future research. The linear quantile mixed effects regression presented in this paper constitutes a practical statistical tool for the evaluation of quantiles with dependent data that complements the widely popular mixed effects regression for the mean and may help to further understanding in many fields of applied research.

Acknowledgements The Centre for Paediatric Epidemiology and Biostatistics benefits from funding support from the Medical Research Council in its capacity as the MRC Centre of Epidemiology for Child Health (G0400546). The UCL Institute of Child Health receives a proportion of funding from the Department of Health's NIHR Biomedical Research Centres funding scheme.

References

- Alhamzawi, R., Yu, K., Pan, J.: Prior elicitation in Bayesian quantile regression for longitudinal data. *J. Biometr. Biostat.* **2**, 1–7 (2011)
- Azzalini, A., Capitanio, A.: Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t -distribution. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **65**, 367–389 (2003)
- Barrodale, I., Roberts, F.D.K.: An efficient algorithm for discrete l_1 linear approximation with linear constraints. *SIAM J. Numer. Anal.* **15**, 603–611 (1978)
- Bassett, G., Koenker, R.: Asymptotic theory of least absolute error regression. *J. Am. Stat. Assoc.* **73**, 618–622 (1978)
- Boscovich, R.J.: *De Litteraria Expeditione per Pontificiam Ditionem, et Synopsis Amplioris Operis, Ac Habentur Plura Ejus Ex Exemplaria Etiam Sensorum Impressa. Bononiesi Scientiarum et Artum Instituto Atque Academia Commentarii*, vol. IV (1757)
- Bose, A., Chatterjee, S.: Generalized bootstrap for estimators of minimizers of convex functions. *J. Stat. Plan. Inference* **117**, 225–239 (2003)
- Bottai, M., Orsini, N.: A command for Laplace regression. *Stata J.* (2012, in press)
- Bottai, M., Zhang, J.: Laplace regression with censored data. *Biom. J.* **52**, 487–503 (2010)
- Buchinsky, M.: Estimating the asymptotic covariance matrix for quantile regression models. A Monte Carlo study. *J. Econom.* **68**, 303–338 (1995)
- Canay, I.A.: A simple approach to quantile regression for panel data. *Econom. J.* **14**, 368–386 (2011)
- Clarke, F.H.: *Optimization and Nonsmooth Analysis*. SIAM, Philadelphia (1990)
- Demidenko, E.: *Mixed Models. Theory and Applications*. Wiley, Hoboken (2004)
- DerSimonian, R., Laird, N.: Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177–188 (1986)
- Doksum, K.: Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Stat.* **2**, 267–277 (1974)
- Eltoft, T., Kim, T., Lee, T.-W.: On the multivariate Laplace distribution. *IEEE Signal Process. Lett.* **13**, 300–303 (2006)
- Farcomeni, A.: Quantile regression for longitudinal data based on latent Markov subject-specific parameters. *Stat. Comput.* **22**, 141–152 (2012)
- Feng, X., He, X., Hu, J.: Wild bootstrap for quantile regression. *Biometrika* **98**, 995–999 (2011)
- Fielding, A., Yang, M., Goldstein, H.: Multilevel ordinal models for examination grades. *Stat. Model.* **3**, 127–153 (2003)
- Fu, L., Wang, Y.-G.: Quantile regression for longitudinal data with a working correlation model. *Comput. Stat. Data Anal.* **56**, 2526–2538 (2012)
- Galvao, A.F.: Quantile regression for dynamic panel data with fixed effects. *J. Econom.* **164**, 142–157 (2011)
- Galvao, A.F., Montes-Rojas, G.V.: Penalized quantile regression for dynamic panel data. *J. Stat. Plan. Inference* **140**, 3476–3497 (2010)
- Genz, A., Keister, B.D.: Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weight. *J. Comput. Appl. Math.* **71**, 299–309 (1996)
- Geraci, M.: lqmm: Linear quantile mixed models. R package version 1.02 (2012)
- Geraci, M., Bottai, M.: Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics* **8**, 140–154 (2007)
- Geraci, M., Salvati, N.: The geographical distribution of the consumption expenditure in Ecuador: estimation and mapping of the regression quantiles. *Stat. Appl.* **19**, 167–183 (2007)
- He, X.: Quantile curves without crossing. *Am. Stat.* **51**, 186–192 (1997)
- He, X., Hu, F.: Markov chain marginal bootstrap. *J. Am. Stat. Assoc.* **97**, 783–795 (2002)
- He, X.M., Ng, P., Portnoy, S.: Bivariate quantile smoothing splines. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **60**, 537–550 (1998)
- He, X.M., Portnoy, S.: Some asymptotic results on bivariate quantile splines. *J. Stat. Plan. Inference* **91**, 341–349 (2000)
- Heiss, F., Winschel, V.: Likelihood approximation by numerical integration on sparse grids. *J. Econom.* **144**, 62–80 (2008)
- Higham, N.: Computing the nearest correlation matrix—a problem from finance. *IMA J. Numer. Anal.* **22**, 329–343 (2002)
- Hinkley, D.V., Revankar, N.S.: Estimation of the Pareto law from underreported data: a further analysis. *J. Econom.* **5**, 1–11 (1977)
- Karlsson, A.: Nonlinear quantile regression estimation of longitudinal data. *Commun. Stat., Simul. Comput.* **37**, 114–131 (2008)
- Kim, M.-O., Yang, Y.: Semiparametric approach to a random effects quantile regression model. *J. Am. Stat. Assoc.* **106**, 1405–1417 (2011)
- Kocherginsky, M., He, X., Mu, Y.: Practical confidence intervals for regression quantiles. *J. Comput. Graph. Stat.* **14**, 41–55 (2005)
- Koenker, R.: Quantile regression for longitudinal data. *J. Multivar. Anal.* **91**, 74–89 (2004)
- Koenker, R.: *Quantile Regression*. Cambridge University Press, New York (2005)
- Koenker, R., Bassett, G.: Regression quantiles. *Econometrica* **46**, 33–50 (1978)

- Koenker, R., Machado, J.A.F.: Goodness of fit and related inference processes for quantile regression. *J. Am. Stat. Assoc.* **94**, 1296–1310 (1999)
- Koenker, R., Mizera, I.: Penalized triograms: total variation regularization for bivariate smoothing. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **66**, 145–163 (2004)
- Koenker, R., Ng, P., Portnoy, S.: Quantile smoothing splines. *Biometrika* **81**, 673–680 (1994)
- Koenker, R., Xiao, Z.J.: Inference on the quantile regression process. *Econometrica* **70**, 1583–1612 (2002)
- Kotz, S., Kozubowski, T.J., Podgórski, K.: An asymmetric multivariate Laplace distribution. Tech. Rep. 367, Department of Statistics and Applied Probability, University of California at Santa Barbara (2000)
- Kozubowski, T.J., Nadarajah, S.: Multitude of Laplace distributions. *Stat. Pap.* **51**, 127–148 (2010)
- Lamarche, C.: Robust penalized quantile regression estimation for panel data. *J. Econom.* **157**, 396–498 (2010)
- Lee, D., Neocleous, T.: Bayesian quantile regression for count data with application to environmental epidemiology. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **59**, 905–920 (2010)
- Lee, Y., Nelder, J.A.: Conditional and marginal models: another view. *Stat. Sci.* **19**, 219–228 (2004)
- Lehmann, E.L.: *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco (1975)
- Li, Q., Xi, R., Lin, N.: Bayesian regularized quantile regression. *Bayesian Anal.* **5**, 533–556 (2010)
- Lipsitz, S.R., Fitzmaurice, G.M., Molenberghs, G., Zhao, L.P.: Quantile regression methods for longitudinal data with drop-outs: application to CD4 cell counts of patients infected with the human immunodeficiency virus. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **46**, 463–476 (1997)
- Liu, Y., Bottai, M.: Mixed-effects models for conditional quantiles with longitudinal data. *Int. J. Biostat.* **5**, 1–22 (2009)
- Lum, K., Gelfand, A.: Spatial quantile multiple regression using the asymmetric Laplace process. *Bayesian Anal.* **7**, 235–258 (2012)
- Machado, J.A.F., Santos Silva, J.M.C.: Quantiles for counts. *J. Am. Stat. Assoc.* **100**, 1226–1237 (2005)
- Oberhofer, W., Haupt, H.: The asymptotic distribution of the unconditional quantile estimator under dependence. *Stat. Probab. Lett.* **73**, 243–250 (2005)
- Parzen, M., Wei, L., Ying, Z.: A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–350 (1994)
- Pinheiro, J., Bates, D.: Approximations to the log-likelihood function in the nonlinear mixed-effects model. *J. Comput. Graph. Stat.* **4**, 12–35 (1995)
- Pinheiro, J.C., Bates, D.M.: Unconstrained parametrizations for variance-covariance matrices. *Stat. Comput.* **6**, 289–296 (1996)
- Pinheiro, J.C., Chao, E.C.: Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. *J. Comput. Graph. Stat.* **15**, 58–81 (2006)
- Pourahmadi, M.: Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika* **86**, 677–690 (1999)
- Prékopa, A.: Logarithmic concave measures and functions. *Acta Sci. Math.* **34**, 334–343 (1973)
- R Development Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna (2012). ISBN 3-900051-07-0
- Reed, W.: The normal-Laplace distribution and its relatives. In: Balakrishnan, N., Castillo, E., Sarabia Alegria, J.-M. (eds.) *Advances in Distribution Theory, Order Statistics, and Inference*, pp. 61–74. Birkhäuser Boston, New York (2006)
- Reich, B.J., Bondell, H.D., Wang, H.J.: Flexible Bayesian quantile regression for independent and clustered data. *Biostatistics* **11**, 337–352 (2010a)
- Reich, B.J., Fuentes, M., Dunson, D.B.: Bayesian spatial quantile regression. *J. Am. Stat. Assoc.* (2010b)
- Rigby, R., Stasinopoulos, D.: Generalized additive models for location, scale and shape. *J. R. Stat. Soc., Ser. C, Appl. Stat.* **54**, 507–554 (2005)
- Robinson, G.: That BLUP is a good thing: the estimation of random effects. *Stat. Sci.* **6**, 15–32 (1991)
- Rockafellar, R.: *Convex Analysis*. Princeton University Press, Princeton (1970)
- Rogan, W.J., Dietrich, K.N., Ware, J.H., Dockery, D.W., Salganik, M., Radcliffe, J., Jones, R.L., Ragan, N.B., Chisolm, J.J., Rhoads, G.G.: The effect of chelation therapy with succimer on neuropsychological development in children exposed to lead. *N. Engl. J. Med.* **344**, 1421–1426 (2001)
- Ruppert, D., Wand, M., Carroll, R.: *Semiparametric Regression*. Cambridge University Press, New York (2003)
- Sarkar, D.: *Lattice: Multivariate Data Visualization with R*. Springer, New York (2008)
- Treatment of Lead-Exposed Children (TLC) Trial Group: Safety and efficacy of succimer in toddlers with blood lead levels of 20–44 µg/dL. *Pediatr. Res.* **48**, 593–599 (2000)
- Wagner, H.M.: Linear programming techniques for regression analysis. *J. Am. Stat. Assoc.* **54**, 206–212 (1959)
- Wang, J.: Bayesian quantile regression for parametric nonlinear mixed effects models. *Stat. Methods Appl.* (2012)
- Yu, K., Lu, Z., Stander, J.: Quantile regression: applications and current research areas. *Statistician* **52**, 331–350 (2003)
- Yu, K., Zhang, J.: A three-parameter asymmetric Laplace distribution and its extension. *Commun. Stat., Theory Methods* **34**, 1867–1879 (2005)
- Yu, K.M., Moyeed, R.A.: Bayesian quantile regression. *Stat. Probab. Lett.* **54**, 437–447 (2001)
- Yuan, Y., Yin, G.: Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics* **66**, 105–114 (2010)
- Zhao, Q.S.: Restricted regression quantiles. *J. Multivar. Anal.* **72**, 78–99 (2000)